Genetic architecture of grain yield phenotypic plasticity in winter wheat

Nicolas Giordano

## Highlight

Please provide a statement that, in fewer than 30 words, highlights the novelty of the paper for the non-expert. The highlight should contain the central findings of your work, as well as keywords and phrases, but should not simply repeat the title. For reviews, the highlight should state the primary objective of the review.

# Abstract

The abstract should be an engaging and informative 'stand-alone' text (without references) of no more than 200 words. Abstracts for research papers should detail why the research was undertaken; the approach and methodology if appropriate; the main findings; and the key conclusions, including wider implications. Repeat keywords and phrases as appropriate. Abbreviations should only be defined and used within the abstract if they are used three times or more in the abstract itself.

# Keywords and Abbreviations

Please list 6–10 keywords (in alphabetical order) after the abstract. Think of words people might use in searches. The main keywords should also appear in the title, abstract, highlight, and subheadings. Natural, consistent repetition of keywords will aid search engine optimisation so others can more easily find and cite the paper.

Please limit the use of non-standard abbreviations, which can make the text harder to read. Generally, only use them for words mentioned more than three times in the text but only where the shortened form will aid readability. Spell out the term on first mention: for example, ‘the International Rice Research Institute (IRRI) is developing rice varieties…’. If you do have such defined abbreviations, also list them, in alphabetical order, after the keywords.

There are a number of common abbreviations that may be used without definition. The same applies to standard chemical symbols.

Authors are asked to avoid using 'CK' as an abbreviation for 'control' as this abbreviation is more commonly used for 'cytokinin' within plant sciences.

# Introduction

This differential performance of cultivars across a given range of environments is due to environmental (E) as well as genotype × environment interaction (G×E) and can be quantified in a trait called phenotypic plasticity.

When breeders release a new cultivar, they expect the variety to have consistently high yields across a wide range of environments. In wheat, plasticity can be a positive or negative trait depending on the source of environmental variation. If phenotypic plasticity of yield is a positive trait, increases in plasticity associate with greater yield potential while maintaining or increasing yield under stress. If phenotypic plasticity of yield is a negative trait, increases in plasticity associate with decreased yield and stress while sustaining or decreasing yield under high yielding conditions. Quantifying how yield plasticity relates to wheat grain yield in the Great Plains remains uncertain, and is crucial to target selection efforts correctly.

Previous literature remarked that plasticity has its own genetic regulation and can be subjected to selection (Bradshaw, 1965; Laitinen and Nikoloski, 2019; Diouf et al., 2020; Monforte, 2020). Breeders select for yield phenotypic plasticity over a minor number of cultivars on the latter generations of the breeding cycle, by testing lines on target population environments. However, in early generations the only way to select for genotypes with the ability to couple with environmental variation would be only on presence of molecular markers. Thus, is crucial to discover quantitative trait loci (QTL) associated with the ability of cultivars to consistently yield high across diverse environments.

In this study we compiled data from variety testing networks on four major wheat producing regions worldwide to answer the following question:

* How has yield phenotypic plasticity changed over time across major wheat producing regions with contrasting germplasms and variety release standards?

In addition we collected QTL data to decipher the genetic architecture of wheat yield phenotypic plasticity and assess whether yield plasticity is under the same genetic regulation as yield per se. We hypothesize that yield plasticity are unrelated and controlled by different genomic region. Furthermore we hypothesize that yield plasticity is control by non-genic regions, such as promoters that trigger the action of genes in certain environmental conditions.

# Materials and methods

## Variety testing network

Data was obtained from the variety testing network in four countries (Argentina, Canada, United States and Uruguay) over 23 years, between 2000 and 2022. Details of the field trials are presented in Table XX and Figure XXX.

## Yield phenotypic plasticity

Grain yield phenotypic plasticity was calculated as the ratio of variances (Eq. XX), being the quotient between the variance of a given genotype and the variance of a population of genotypes across environments (Lin et al., 1986; Giordano et al., 2024). The variance ratio gives a genotype-specific estimate for plasticity and is comparable with the slope of a reaction norm of yield against yield environment (Sadras and Richards, 2014, Figure XXX). The variance ratio does not require any assumptions on the process model behind the estimation of yield plasticity. To avoid sample size bias in the calculation of yield plasticity, we included only cultivars that were tested in at least 25 environments (Table XXX).

## Genotyping

For genotypes present in at least 25 environments within the Central Great Plains region, DNA was extracted using XXX machine from manufacturer. Genotyping by sequence was done using TASSEL5-Reference pipeline was used with IWGSC\_RefSeqv2.1 reference genome to find markers in the DNA samples. Single nucleotide polymorphisms (SNPs) with <= 20% missing markers, and >=5 % minor allele frequency were retained. The final number of SNPs was 18,789.

## Data analysis

Observed grain yield was modeled as:

Here, is the yield observation with expected value which is linked to the linear predictor through a log-link function, while represents the dispersion. Thus, the process model for includes which is an matrix of genotypes (), while is a vector of length of random effects for or BLUPs. Similarly, is a matrix of size representing the incidence of each environment (). is a vector of length containing the random effects. Finally, is a matrix of size for random effects and a vector of length XXXX for the random effects. We run a gaussian model, compare it against the gamma and retained the one with the lowest Bayesian Information Criterion (BIC).

To explore the genetic changes on traits of interest over time we modeled the association between the genotype year of release and (i) grain yield best linear unbiased predictors (BLUPs, ) and (ii) grain yield plasticity.

## Trait changes over time

To explore the genetic trends over time of grain yield per se and yield plasticity we fitted for the expected trait value (Eq XX and Eq XX) as follow:

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

where represents the genotype-specific yield BLUPs () or yield plasticity () within a given region of interest. The expected value of the response variable is which is linked to the linear predictor through a log-link function, while represents the dispersion.

In Argentina, we modeled 90th percentile of yield plasticity changes with year of release of the genotype as:

|  |  |
| --- | --- |
|  |  |
|  |  |

In Eq. XX we assumed yield plasticity in Argentina follows an asymmetric Laplace distribution where the expected value of is equal to and has a dispersion equal to .

Similarly, we explored the association of yield *per se* and yield plasticity with an asymmetric Laplace quantile regression model (Eq. XX). We modeled both 10th and 90th percentiles to make inference on this association at high and low yielding environments, respectively.

|  |  |
| --- | --- |
|  |  |
|  |  |

Furthermore, to model the association of yield BLUPs and yield plasticity we fitted two process models for the expected value (, Eq. XX, XX) and selected the one with lowest BIC.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

Models in Eq. XX, XX, XX were fitted using the *glmer* function available in R. Models in Equations XX, XX, XX were fitted using the *brms* package in R. Details of Bayesian models fitted and priors used are provided in Appendix 1. All bayesian models were fitted using uninformative priors.

## Genome wide association study

# Results

# Discussion